Published in final edited form as:

Mayo Clin Proc. 2015 September ; 90(9): 1171–1175. doi:10.1016/j.mayocp.2015.05.006.

Hard-Wired Bias: How Even Double-Blind Randomized Controlled Trials Can Be Skewed from the Start

Vinay Prasad, MD, MPH and

Medical Oncology Branch, National Cancer Institute, National Institutes of Health, 10 Center Dr. 10/12N226, Bethesda, MD 20892, Phone: (219) 22900170; Fax (301) 402-1608

Vance W. Berger, PhD

National Cancer Institute and University of Maryland Baltimore County, Biometry Research Group, Division of Cancer Prevention, National Cancer Institute, National Institutes of Health, 9609 Medical Center Drive, Rockville, MD 20850, (240) 276-7142 (voice)

Vinay Prasad: vinayak.prasad@nih.gov; Vance W. Berger: vb78c@nih.gov

Well designed, adequately powered randomized controlled trials (RCTs) are rightfully considered the highest form of evidence on which to base treatment and diagnostic decisions, minimizing potential biases, particularly confounding, that plague alternate, lesser forms of evidence[1–3]. At the same time, simply being an RCT is not sufficient to ensure that conclusions are free from bias. In recent years, sponsors and trialists have incorporated subtle design choices into RCTs that have skewed the final trial results. Here, we describe the phenomenon of "hard wired" bias—bias that is introduced at the outset of randomized trials[4]. We hope that recognition of hard-wired bias serves as a reminder that there remains room for improvement in the analysis and conduct of clinical trials.

Bias in the interpretation of clinical data may occur at many junctures, as depicted in, but not limited to, the Figure. Selective reporting of outcomes, and publication bias occur at the final step in the presentation of data. Bias in the interpretation and analysis of data, including deviations from preplanned statistical analysis occur penultimately. Bias can also occur *prior* to the collection of data—implicit in the design of clinical trials. Unlike bias in the analysis or reporting of data, bias in trial design cannot be overcome by statistical methods or reanalysis[5]; instead it can only be noted as a limitation of the study. What are some of the ways in which bias in clinical trials can become hard-wired?

Address correspondence to: Vinay Prasad, MD MPH, National Cancer Institute, National Institutes of Health, 10 Center Dr. 10/12N226, Bethesda, MD 20892, Phone: (219) 22900170; Fax (301) 402-1608, vinayak.prasad@nih.gov.

Website for NY Times, Krumholz: http://www.nytimes.com/2014/02/03/opinion/give-the-data-to-the-people.html
Website for NICE: http://www.nice.org.uk/guidance/TA219/chapter/4-Consideration-of-the-evidence
Website for AHRQ: http://www.cms.gov/Medicare/Coverage/DeterminationProcess/downloads/id77ta.pdf

Crossover

Crossover from placebo or control to the investigational agent is typically done in pharmacokinetic studies, or for interventions aimed at assessing a subjective endpoint (permitting intra-user comparisons). However, in modern clinical trials, crossover is increasingly used in studies testing the *basic efficacy* of a novel compound. For instance, many randomized controlled trials of cancer therapeutics allow patients assigned to placebo to crossover to treatment upon progression of their disease[6]. Such crossover affects interpretations regarding the drug's effect on survival. For instance, if a cancer drug delays progression, but does not improve survival, crossover is often cited as the reason for these findings. *The drug would have improved survival had it not been for crossover*. However, despite this common interpretation, there are other valid interpretations in this setting. For instance, a drug may slow progression, but increase off-target deaths, such that it has no net benefit[7]. In this case, crossover can mask the harms of the medication, and provide a misleading inference regarding benefits.

Sipuleucil T (Provenge, Dendreon) is a cancer vaccine approved by the US Food and Drug Administration (FDA) for the treatment of metastatic prostate cancer. In the seminal trial leading to approval, the drug managed to improve overall survival without any evidence that it slowed disease progression[8]. In the trial, many patients in the control arm received the frozen vaccine when their cancer progressed, and fewer patients received docetaxel—a drug with a proven survival benefit—or received it after a delay. This design led to the suggestion that Sipuleucel-T demonstrated efficacy not by improving outcomes, but rather because crossover harmed the control group—by delaying alternate, effective therapy[9]. When it comes to crossover, better conclusions cannot be obtained simply through access to the data —all interpretations of the data have to make some assumptions about whether crossover is benefitting or harming the control group. Consider RECORD-1, a randomized trial comparing everolimus to placebo in patients with metastatic renal cell cancer for whom prior therapy had failed. RECORD-1 demonstrated an improvement in progression free survival, but no change in overall survival[10]. This deficiency was attributed to crossover[10], and the manufacturer provided modeling experiments arguing what the survival would have been were it not for crossover[11]. However, this exercise relies on assumptions, which may not be true. Because we cannot know the true effect of everolimus on survival, it was rejected by the UK's National Institute for Clinical Excellence (NICE).[12]

For a different set of questions however, a lack of crossover can bias results. When clinical trials seek to establish the basic efficacy of a novel compound, the presence of crossover can distort inferences, as we have shown. However, when clinical trials test the proper sequencing of agents *already known to confer benefit*, the *absence* of crossover can be equally problematic. Consider some contemporary examples. In a randomized trial from Spain that showed lenalidomide and dexamethasone improved overall survival among patients with smoldering myeloma over current standard of care of observation[13], patients in the control arm were not routinely given access to lenalidomide when they developed overt multiple myeloma. Thus we cannot be sure that the survival advantage of early treatment would still exist if control patients had fair access to this drug, as they would have had in the US. The failure to proscribe lenalidomide upon progression is also a limitation of randomized trials of maintenance therapy[14,15], which (should) ask the question of whether continuous administration of a drug is better than receiving the agent upon having progressive disease.

In a final example, the proteasome inhibitor bortezomib, which is FDA approved for relapsed mantle cell lymphoma, was tested in the first line setting of this disease. At the time of the trial, the drug was widely used in the relapsed setting. Thus, any trial seeking to advance the drug into the first line should show that early use of the drug improves survival beyond its second line use. However, the randomized trial testing this question was globally conducted, and, as such, patients in the control arm had poor access to the drug in the second line setting, with only 19% receiving it[16]. Thus, we cannot be sure that the survival advantage seen in the trial would exist had it been conducted in nations where bortezomib is a mainstay of second line treatment.

Selection Bias

Bias in choice of study subjects is a frequent concern in clinical trials. Selection bias can be caused by the inclusion and exclusion criteria to a study, which prevent the generalizability of results to a broader patient population. Although such concerns are increasingly appreciated, there remain pertinent even today. For example, among cancer drugs approved by US Food and Drug Administration between 1995 and 2002, demographics of patients were strikingly different from cancer patients in the United States[17]. While the proportions of patients aged ≥ 65, ≥ 70, and ≥ 75 years was 60%, 46%, and 31%, respectively among cancer patients in the United States, these age groups comprised only 36%, 20%, and 9% of patients in registration trials ($P < .001$). Fehrenbacher and colleagues extended these findings, and showed that inclusion criteria of contemporary randomized trials in non-small cell lung cancer would exclude the majority of patients treated at Kaiser Permanente, a large insurer with a representative patient population[18].

Although these examples are illustrative, this problem is relatively tractable, as care can be taken to prevent extrapolation to untested populations. However, in other cases, selection bias cannot be accounted for.

Consider the open-label “run in period.” The Heart Protection Study, a randomized trial of 20,536 high cardiovascular risk individuals, tested whether simvastatin at 40mg daily could improve outcomes compared to placebo[19]. The trial found that the medication decreased major vascular events by 25%, and the authors go further, arguing that, without non-compliance, the improvement would have been 33%[19]. Notably, this trial utilized a 4-week placebo run in period followed by a 4–6 week simvastatin run in period prior to randomization. During this time, a patient’s primary doctor could remove a patient from randomization, and any patient could elect not to be randomized for “any reason[20].” All together, 11,609 patients who were eligible for the study and began the run-in period dropped out prior to randomization[20]. Thus, over a third of the patients who began the study were not randomized, and no set of specified inclusion criteria can define the set of patients who remained. Others have noted that the use of run-in periods can limit the applicability of study findings, and can inflate estimates of benefits[21]. This occurs in part because run-in periods of the active drug tests a different clinical question—whether discontinuation of a therapy is harmful—rather than whether initiating a therapy is beneficial.

Open-label run in periods were also problematic in the PARADIGM-HF trial[22]. This study randomized 8442 patients predominantly with NY Heart II and III heart failure to a combination of valsartan (an angiotensin receptor blocker (ARB)) and neprylisin (the investigational agent) or enalapril (the control arm). Yet, prior to randomization, over 10,500 patients entered into a run in period. During this time, patients were sequentially treated with a median of 15 days of enalapril followed by 29 days of the combination medication. Nearly 20% of participants who began the study dropped out of the study during this time. Thus, the run in period created both an indefinable study population, viz. patients who met inclusion criteria and did not drop out after 15 days of enalapril followed by 29 days of the combination medication, as well as posed a different question, whether switching back to enalapril was better or worse than continuing the combination among such patients.

The Unlevel Playing Field

Another way modern trials have hard-wired bias is by promoting an unequal comparison. For instance, a head to head trial of the tyrosine kinase inhibitors axitinib and sorafenib in metastatic kidney cancer appears to be fair; however a closer examination reveals problems[23]. Specifically, while the starting dose of both drugs is appropriate, the dose reductions for toxicity favor the axitinib arm. For similar side effects, sorafenib has steeper dose reductions, and for patients doing well on full dose axitinib, the dose could even be *increased.* Collectively this meant that axitinib was pushed to a higher dose, and penalized less than sorafenib[24]. It should be no surprise which medication was declared the winner of that study.

Another type of unequal playing field occurs when two cancer drugs are tested head to head, but as a matter of fact more patients had already taken, and had the opportunity to have their cancer acquire resistance to, one of those drugs. In the ENDEAVOR randomized trial, patients with multiple myeloma who had had their disease progress on treatment were randomized to bortezomib or carfilzomib in combination with dexamethasone[25]. The results showed a progression free survival benefit for carfilzomib, but as a matter of fact, given the date these drugs were approved, the majority of patients had the opportunity to be previously treated with bortezomib, while very few patients had the chance to previously receive carfilzomib. As a general rule in cancer, two drugs can be on average comparable, but each is less effective among patients already treated with that medication—a bias that the ENDEAVOR trial exploits.

Control Arms

The control for a clinical trial is ideally selected depending on the clinical question posed. Controls should reflect the best therapy currently being used in the target population, and for studies evaluating subjective endpoints, the control should be *as close as possible* to the investigational arm. If this condition is not met, a trial essentially uses a straw man comparator. Many times, the use of a sham control has unmasked bias in studies supporting the use of a medical procedure. For instance, vertebroplasty, epidural steroid injection, and arthroscopic meniscectomy all required a sham-controlled trial to demonstrate that the treatments had no benefit[26].

Censoring

Informative censoring[27] in clinical trials can distort our perception of the benefits of a treatment. All survival analysis is based on the premise that censoring is uninformative—the patients censored are no different than those who are followed. However, increasingly this assumption may be questioned. In many cancer treatment trials, censoring often occurs as patients withdraw from toxicity or intolerability. These patients are likely to be different from those who tolerate therapy well. In a recent study, Campigotto and Weller provide two examples in which patients who are censored are likely to have better or worse survival than those who remain on study[28]. The authors then provide a range of estimates for the outcome had these patients not been excluded, based upon simulation. But, if patients come off treatment, and are no longer followed, we cannot reconstruct their outcomes without making assumptions. We are left with a hard wired bias.

Conclusion

Randomized control trials remain the best way to draw sound conclusions regarding the efficacy and impact of drugs, devices, screening and diagnostic tests, but unfortunately randomization does not ensure a fair trial. In this respect, randomization can be viewed as necessary but not sufficient for sound scientific decision making. The purpose of our analysis is not to disparage the growth of RCTs—which we believe is inevitable and unquestionably valuable—but to highlight persistent challenges. Many types of bias can be remedied by access to individual patient level data, while other types of bias, so called “hard-wired” bias, cannot be corrected for after the fact.

Medical trials involve the participation of human subjects, who donate their time and energy to further the altruistic pursuit of improved medical care. As such, we have a moral obligation to ensure that research is capable of most honestly answering an important clinical question. The elements of trial design that we discussed—crossover, drug run-in periods, the use of inadequate controls, early censoring, selection bias, and duration of follow up—are decisions made at the outset of a clinical trial, and cannot be later corrected. We must work to remove hard-wired bias from clinical trials, and only time will tell if our existing system (in which trials are predominantly funded and conducted by industry) can meet this challenge.

Acknowledgments

The views and opinions of the authors do not reflect those of the National Cancer Institute We are both US Federal employees and copyright cannot be transferred.

References

1. Guyatt G. EVIDENCE-BASED MEDICINE - A NEW APPROACH TO TEACHING THE PRACTICE OF MEDICINE. Jama-Journal of the American Medical Association. 1992; 268:2420–5.

2. Ioannidis JP. Why most published research findings are false. PLoS medicine. 2005; 2:e124. [PubMed: 16060722]
3. Prasad V, Jorgenson J, Ioannidis JP, Cifu A. Observational studies often make clinical practice recommendations: an empirical evaluation of authors' attitudes. Journal of clinical epidemiology. 2013; 66:361–6. e4. [PubMed: 23384591]
4. Senn, HJKU.; Otto, F. Cancer Prevention II (Recent Results in Cancer Research). Berlin: Springer; 2008.
5. Ebrahim S, Sohani ZN, Montoya L, et al. REanalyses of randomized clinical trial data. JAMA: the journal of the American Medical Association. 2014; 312:1024–32. [PubMed: 25203082]
6. Prasad V, Grady C. The misguided ethics of crossover trials. Contemporary clinical trials. 2014; 37:167–9. [PubMed: 24365533]
7. Prasad V. Double-crossed: why crossover in clinical trials may be distorting medical science. Journal of the National Comprehensive Cancer Network: JNCCN. 2013; 11:625–7. [PubMed: 23667211]
8. Di Lorenzo G, Ferro M, Buonerba C. Sipuleucel-T (Provenge(R)) for castration-resistant prostate cancer. BJU international. 2012; 110:E99–104. [PubMed: 22177289]
9. [Accessed March 6, 2014] Outcomes of Sipuleucel-T therapy. at http://www.cms.gov/Medicare/Coverage/DeterminationProcess/downloads/id77TA.pdf
10. Motzer RJ, Escudier B, Oudard S, et al. Efficacy of everolimus in advanced renal cell carcinoma: a double-blind, randomised, placebo-controlled phase III trial. Lancet. 2008; 372:449–56. [PubMed: 18653228]
11. Korhonen P, Zuber E, Branson M, et al. Correcting overall survival for the impact of crossover via a rank-preserving structural failure time (RPSFT) model in the RECORD-1 trial of everolimus in metastatic renal-cell carcinoma. Journal of biopharmaceutical statistics. 2012; 22:1258–71. [PubMed: 23075021]
12. [Accessed 8-30-2014] Everolimus for the second-line treatment of advanced renal cell carcinoma. 2011. at http://www.nice.org.uk/guidance/TA219/chapter/4-Consideration-of-the-evidence
13. Mateos M-V, Hernández M-T, Giraldo P, et al. Lenalidomide plus Dexamethasone for High-Risk Smoldering Multiple Myeloma. New England Journal of Medicine. 2013; 369:438–47. [PubMed: 23902483]
14. Attal M, Lauwers-Cances V, Marit G, et al. Lenalidomide maintenance after stem-cell transplantation for multiple myeloma. The New England journal of medicine. 2012; 366:1782–91. [PubMed: 22571202]
15. Palumbo A, Cavallo F, Gay F, et al. Autologous transplantation and maintenance therapy in multiple myeloma. The New England journal of medicine. 2014; 371:895–905. [PubMed: 25184862]
16. Robak T, Huang H, Jin J, et al. Bortezomib-Based Therapy for Newly Diagnosed Mantle-Cell Lymphoma. New England Journal of Medicine. 2015; 372:944–53. [PubMed: 25738670]
17. Talarico L, Chen G, Pazdur R. Enrollment of Elderly Patients in Clinical Trials for Cancer Drug Registration: A 7-Year Experience by the US Food and Drug Administration. Journal of Clinical Oncology. 2004; 22:4626–31. [PubMed: 15542812]
18. Fehrenbacher LAL, Somkin C. Randomized clinical trial eligibility rates for chemotherapy (CT) and antiangiogenic therapy (AAT) in a population-based cohort of newly diagnosed non-small cell lung cancer (NSCLC) patients. J Clin Oncol. 2009; 27(suppl):15s. abstr 6538.
19. MRC/BHF Heart Protection Study of cholesterol lowering with simvastatin in 20,536 high-risk individuals: a randomised placebo-controlled trial. Lancet. 2002; 360:7–22. [PubMed: 12114036]
20. Armitage J, Bowman L, Collins R, Parish S, Tobert J. Effects of simvastatin 40 mg daily on muscle and liver adverse effects in a 5-year randomized placebo-controlled trial in 20,536 high-risk people. BMC clinical pharmacology. 2009; 9:6. [PubMed: 19442259]
21. Pablos-Mendez A, Barr RG, Shea S. Run-in periods in randomized trials: implications for the application of results in clinical practice. JAMA: the journal of the American Medical Association. 1998; 279:222–5. [PubMed: 9438743]
22. McMurray JJV, Packer M, Desai AS, et al. Angiotensin–Neprilysin Inhibition versus Enalapril in Heart Failure. New England Journal of Medicine. 2014; 371:993–1004. [PubMed: 25176015]

23. Rini BI, Escudier B, Tomczak P, et al. Comparative effectiveness of axitinib versus sorafenib in advanced renal cell carcinoma (AXIS): a randomised phase 3 trial. Lancet. 2011; 378:1931–9. [PubMed: 22056247]

24. Prasad V, Massey PR, Fojo T. Oral Anticancer Drugs: How Limited Dosing Options and Dose Reductions May Affect Outcomes in Comparative Trials and Efficacy in Patients. Journal of Clinical Oncology. 2014

25. [Accessed 5-3-2015] Carfilzomib Doubles PFS Over Bortezomib in Phase III Multiple Myeloma Trial. 2015. at http://www.onclive.com/web-exclusives/Carfilzomib-Doubles-PFS-Over-Bortezomib-in-Phase-III-Multiple-Myeloma-Trial

26. Redberg RF. Sham Controls in Medical Device Trials. New England Journal of Medicine. 2014; 371:892–3. [PubMed: 25184861]

27. Uno H, Claggett B, Tian L, et al. Moving Beyond the Hazard Ratio in Quantifying the Between-Group Difference in Survival Analysis. Journal of Clinical Oncology. 2014; 32:2380–5. [PubMed: 24982461]

28. Campigotto F, Weller E. Impact of Informative Censoring on the Kaplan-Meier Estimate of Progression-Free Survival in Phase II Clinical Trials. Journal of clinical oncology: official journal of the American Society of Clinical Oncology. 2014

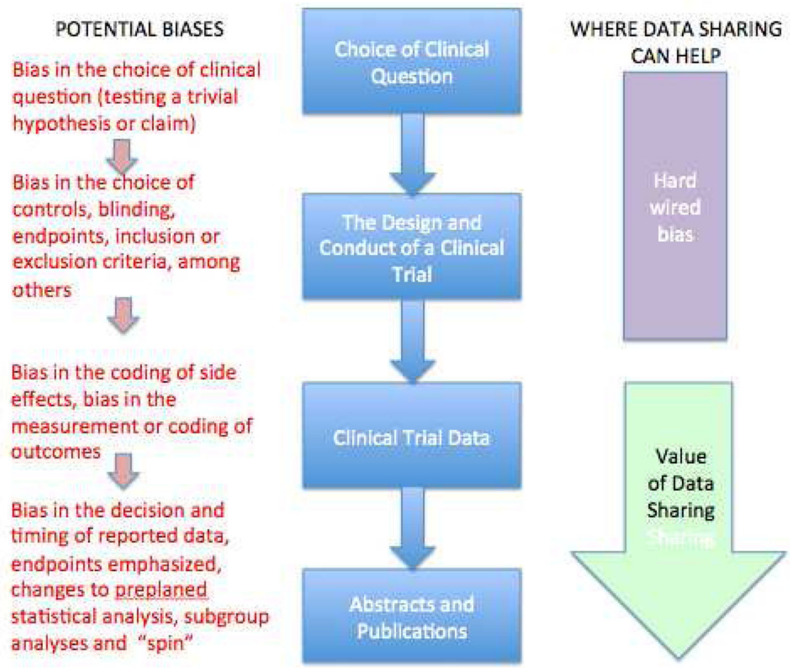

Figure 1.
The origins of bias in clinical trials